DOCUMENT RESUME

ED 316 584 TM 014 561

AUTHOR Tucker, Mary L.
TITLE A Compendium of Textbook Views on Planned versus Post Hoc Tests.
PUB DATE 25 Jan 90
NOTE 30p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 25-27, 1990).
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Variance; Graduate Study; Higher Education; Literature Reviews; *Research Methodology; *Statistical Significance; *Textbook Content
IDENTIFIERS *A Priori Tests; Multiple Comparisons; *Post Hoc Methods; Power (Statistics)

ABSTRACT

This paper focuses on two methods of performing multiple comparisons (post hoc or unplanned versus a priori or planned) for determining exactly where statistically significant results arise when there are more than "k equals two" groups. Textbook views and recommendations regarding the use of these two types of multiple comparisons are summarized. Analysis of variance (ANOVA) and related methods remain popular, although other methods are sometimes more useful and although the frequency of use of these methods in published research is declining. Researchers conducting an ANOVA frequently conduct unplanned or post hoc tests to determine exactly where statistically significant group differences actually arise. Various methodologists have referred to this practice as "fishing", "milking", or "data snooping". It is widely known that planned multiple comparisons have greater statistical power than do unplanned multiple comparisons. Secondly, and more importantly, planned comparisons tend to make the researcher think in advance. Graduate study in statistics should reflect the importance of planned versus post hoc statistical design. In addition, there is a need for further investigation of current statistics textbooks being used in the different schools of education, business, and the behavioral sciences to determine if the issue of planned versus post hoc tests is being adequately addressed in instructional materials. A 63-item list of references, 4 data tables, and a list of selected computer control cards are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED316584

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Mary L. Tucker
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A COMPENDIUM OF TEXTBOOK VIEWS ON PLANNED VERSUS POST HOC TESTS

Mary L. Tucker

Nicholls State University

and

Graduate School
University of New Orleans

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 25, 1990.

BEST COPY AVAILABLE

TM014561

Abstract

ANOVA and related methods remain popular, though other methods are sometimes more useful (Thompson, 1986) and though the frequency of use of these methods in published research is declining (Elmore & Woehlke, 1988; Gaither & Glorfeld, 1985). Researchers conducting ANOVA frequently conduct unplanned or post hoc tests to determine exactly where statistically significant group differences actually arise. Various methodologists have referred to this practice as "fishing" (Minium & Clarke, 1982, p. 321), "milking" (Keppel, 1982, p. 150) or "data snooping" (Kirk, 1969, p. 73; 1984, p. 360; Pedhazur, 1982, p. 305). An alternative is to conduct planned or a priori tests rather than post hoc or unplanned tests. It is widely known that planned multiple comparisons have greater power than unplanned multiple comparisons (Thompson, 1985, 1988b, 1988c, 1989). The purpose of this paper is to review and summarize textbook views and recommendations regarding the use of these two types of multiple comparisons (Thompson, 1987). Such a review may be useful in understanding the origins of several features of contemporary analytic practice.

Researchers have several analytical methods for evaluating differences between means. Often each method will provide different results. Since Fisher (1925) introduced the analysis of variance (ANOVA) methods, there has been a preponderance of ANOVA methods and their analogs (ANCOVA, MANOVA, AND MANCOVA)--hereafter referred to as "OVA methods" (Thompson, 1985)--used in educational research. From 1969 to 1978, according to Willson (1980), OVA methods were utilized in 56 percent of the articles published in the American Educational Research Journal (AERJ). Thirty-seven percent of the AERJ articles between 1979-1983 used OVA techniques (Goodwin & Goodwin, 1985a, 1985b). Reviews by Elmore and Woehlke (1988) show that ANOVA and ANCOVA methods make up 25 percent of the analytic methods included in publications of three educational research journals between 1978-1987. Even though the trend painted by these reviewers and by others (Gaither & Glorfeld, 1985) show OVA utilization on the decline, these methods remain popular (Daniel, 1989)--although other methods are sometimes more useful (Thompson, 1986). Several concerns have been voiced by some researchers regarding this continued infatuation with OVA methods. These concerns are beyond the scope of this paper, but are explained elsewhere (Cohen, 1968; Thompson, 1981).

OVA analytical methods examine the null hypothesis of equality among means. If there are more than k = 2 levels in a way or factor and the null hypothesis is rejected, the researcher using classical OVA methods must then conduct additional tests to determine from where the differences

arise (DuRapau, 1988). This is emphasized by Moore (1983, p. 299) when he suggests that:

> If we have statistical significance when we have only two groups, and thus only two means, we can visually inspect the data to determine which group performed better than the other. But when we have three or more groups, we need to investigate specific mean comparisons.

Purpose and Scope

This paper focuses on two methods of performing multiple comparisons (post hoc or unplanned versus a priori or planned) for determining exactly where statistically significant results arise when there are more than k = two groups. The paper summarizes textbook views and recommendations regarding the use of these two types of multiple comparisons (Thompson, 1987). Such a review may be useful in understanding the origins of several features of contemporary analytic practice.

Post Hoc or Unplanned Tests

When a researcher obtains a significant F in OVA analysis and k = 3 or more, there is a question about where the effect is coming from (Gravetter & Wallnau, 1985). Huck, Cormier, and Bounds (1974, p. 68) remind us that "a researcher cannot stop his analysis after getting a significant F." This can be achieved by employing unplanned, post hoc, or a posteriori multiple comparisons tests--such as Sheffe, Tukey, or Duncan--to determine exactly where statistically significant group

differences actually arise within ways (OVA factors) having three or more levels (Thompson, 1988a; 1988b).

A posteriori comparisons may include all possible comparisons and, thus, are seen by some researchers as reflecting a lack of pre-experiment thought and preparation. Post hoc tests also result in the inflation of the experimentwise Type II error rate. Post hoc tests incorporate statistical adjustments to correct for inflation of the experimentwise Type I error (Fish, 1988) associated with conducting so many tests. This causes unplanned tests to have less power in controlling for Type II error.

Huberty and Morris (1988, p. 576) advocate a single contrast test procedure and believe that the distinction between planned and post hoc tests made by some textbook authors is "drastically overplayed." They maintain that "only very few research situations would preclude a researcher from specifying all contrasts of interest prior to an examination of the outcome measures and/or the outcome 'cell' means. [A typical set of contrasts investigated consists of, simply, all pairwise comparisons.]" However, this view is somewhat atypical.

Various methodologists have referred to the use of post hoc testing as "milking" (Keppel, 1982, p. 150) or "data snooping" (Kirk, 1969, p. 73; 1984, p. 360; Pedhazur, 1982, p. 305). In much the same manner, Minium and Clarke (1982, p. 321) state:

> Prior to running the experiment, the investigator in our example had no well-developed rationale for focusing on a

> particular comparison between means. His was a "fishing expedition"... Such comparisons are known as post hoc comparisons, because interest in them is developed "after the fact"--it is stimulated by the results obtained, not by any prior rationale.

Kerlinger (1986, p. 219) concedes that a posteriori comparisons can be useful in research "for exploring one's data and for getting leads for future research," but suggests that a priori comparisons are "perhaps more important scientifically." Wise researchers know that all univariate statistics are a special case of regression and that OVA methods can be utilized using regression approaches to make planned comparisons of selected means.

Planned or A Priori Tests

This alternative to post hoc or unplanned tests is achieved by constructing planned (also called a priori or focused) comparisons using sets of weighting contrasts. It is possible to code the contrasts to test for non-trend hypotheses focusing on selected groups, or the researcher can use trend contrasts to test hypotheses about the patterns in the means across all the groups. Thompson (1987) generated a data set to simulate a non-trend contrast analysis in the six-level, one-way case, shown in Table 1. Contrasts testing for trends in means are provided by Fisher and Yates (1957, pp. 90-100) and by Hicks (1973) and could have been used instead of non-trend contrasts if the interest had been in testing hypotheses about patterns in the group means across all groups.

INSERT TABLE 1 ABOUT HERE

For each main effect, there is one less orthogonal (i.e., uncorrelated) coding contrast than the number of levels in that way. Thus, the total number of coding columns in Table 1 equals the number of degrees of freedom, i.e., df = k - 1 where k = the number of levels in the way. Each coding column in the example depicts a contrast that represents a null hypothesis (DuRapau, 1988, p. 3).

The information involved in testing omnibus OVA effects is the same as that represented in the coding columns as a complete set. The difference in the contrast coding lies in the partitioning of the total explained sum of squares, SOS explained, of the dependent variable into smaller parts using separate one df hypotheses (Thompson, 1985).

Contrasts do not have to be orthogonal. Winer (1971, p. 175) stated that "whether these comparisons are orthogonal or not makes little or no difference." Orthogonal planned comparisons have a special appeal, however. Keppel (1982, p. 147) stated that, since orthogonal comparisons are uncorrelated, decisions regarding the null hypothesis of one comparison (or contrast) is not influenced by decisions regarding any other orthogonal comparison. This independence seems to be an important factor in the analysis of relationships among means. When each contrast (coding column) totals zero and when the sum of the cross-products of each pair of contrasts total zero, contrasts are perfectly uncorrelated or orthogonal, as are the contrasts presented in Table 1 (Thompson, 1988b).

Preference for Planned Comparisons

For two reasons, planned comparisons are generally better than post hoc tests. As stated earlier, the first reason is that planned multiple comparisons have greater statistical power against Type II error than unplanned multiple comparisons (Thompson, 1985; 1988b; 1988c). Power means one is more likely to get statistical significance for fixed sample and effect sizes. Thompson (1987, 1988b) demonstrates that with the Table 1 data. With classical ANOVA there were no statistically significant effects, but significant effects were achieved using planned comparisons with regression and planned comparisons. Thus, Kerlinger and Pedhazur (1973, p. 131) stated:

> The tests of significance for a priori, or planned, comparisons are more powerful than those for post hoc comparisons. In other words, it is possible for a specific comparison to be not significant when tested by post hoc methods but significant when tested by a priori methods.

Similarly, Rosnow and Rosenthal (1989, p. 1281) deplore the "overreliance on omnibus tests of diffuse hypotheses that although providing protection for some investigators from the dangers of 'data mining' with multiple tests performed as if each were the only one considered" because omnibus tests generally do not:

> tell us anything we really want to know. As Abelson (1962) pointed out long ago in the case of analysis of variance

> (ANOVA), the problem is that when the null hypothesis is accepted, it is frequently because of the insensitive omnibus character of the standard F-test as much as by reason of sizable error variance. All the while that a particular predicted pattern among the means is evident to the naked eye the standard F-test is often insufficiently illuminating to reject the null hypothesis that several means are statistically identical.

There are critics of significance testing (Carver, 1978) who would agree with Thompson (1988a, p. 100) that "significance is not, however, the end-all and be-all of research." But there can be no doubt that significance is paramount in research for graduate students:

> It may not be an exaggeration to say that for many PhD students, for whom the .05 alpha has acquired almost an ontological mystique, it can mean joy, a doctoral degree, and a tenure-track position at a major university if their dissertation p is less than .05. However, if the p is greater than .05, it can mean ruin, despair, and their advisor's suddenly thinking of a new control condition that should be run...We only wish to emphasize that dichotomous significance testing has no otological basis. That is, we want to underscore that, surely, God loves the .06 nearly as much as the .05. Can there be any doubt that God views the strength of evidence for or against the null as a fairly

continuous function of the magnitude of p (Rosnow & Rosenthal, 1989, p. 1277).

The second and more substantial reason for choosing to employ a priori contrasts is because planned comparisons tend to make the researcher think--planned comparisons must be drafted prior to data collection and generally only a given number are stated in a study. Usually this is helpful. Keppel (1982, p. 165) spoke of planned comparisons as "the motivating force behind an experiment" and stated that planned comparisons "represent an interest in particular combinations of conditions--not in the overall experiment." Diligently planning the design of the research effort is the foundation for planned comparisons. This is in opposition to post hoc methods that may lead to haphazard research design that waits to focus on particular comparisons between means until achieving a significant omnibus overall F test.

Researchers who feel ANOVA is the best analytic method for their study might get more interpretable, and more correct results by using ANOVA through regression methods with planned comparisons when there are more than two levels in any one way. Researchers might also take notice that there is NEVER a research in which the regression statistical procedure is not at least as good as OVA methods. There is nothing that can be accomplished by OVA methods that cannot be done by regression approaches. An added perk in using regression analyses for data higher than nominal scale is that variance is not squandered. "Among the practical advantages of contrasts," according to Rosnow & Rosenthal, 1989,

p. 1281), "are that they can be easily computed with a pocket calculator, can be computed on the data in published reports as well as with original data, and most important, usually result in increased power and greater clarity of substantive interpretation."

Analysis of Table 1

Table 1 data can by analyzed to concretely illustrate the points made thus far. These contrasts are orthogonal because the five coding columns ($k - 1 = 5$) each total zero and the sums of their cross-products also all equal zero. There is no interaction in this one-way, balanced design, hence no interaction coding columns are developed for these data.

Classical omnibus analysis of variance methods were employed to determine if any differences among the means of the groups existed for the data in Table 1. As shown in Table 2, the omnibus test of differences among the six group means is not statistically significant ($F = 1.5$, $df = 5/6$, $p = .3155$). Additionally, if a post hoc test was run in violation of traditional practice (because the one-way ANOVA did not involve a significant F), there still would not be any statistically significant differences among means for these data.

INSERT TABLE 2 ABOUT HERE

ANOVA through regression is more appropriate with this study because there are six levels to this one-way design and $k > 2$. As illustrated in Table 3, using the regression approach to ANOVA with planned comparisons,

the mean of the two level-six subjects versus the mean of the remaining 10 subjects is statistically significant (F = 12.5, df = 1/6, p = .0054). The null hypothesis, i.e., that the two school board members mean attitude-toward-school score (DV) equals the mean for the other 10 subjects, is rejected (DuRapau, 1988; Thompson, 1987).

INSERT TABLE 3 ABOUT HERE

All the same conclusions apply in the multivariate case, as emphasized by Swaminathan (1989, p. 231):

> The often advocated procedure of following up the rejection of the null hypothesis with a more powerful multiple comparison procedure should be discouraged. First, the overall rejection of the null hypothesis does not guarantee that any meaningful contrast among the means will be significant...Second, since the overall test and the follow up procedure are unrelated, significant contrasts may be found even when the null hypothesis would not have been rejected. Third, follow up multiple comparison procedures which are unrelated to the overall test result in an inflation of the experiment-wise error rate. If multiple comparisons are of primary interest, a suitable multiple comparison procedure can be used without first performing an overall test.

Review and Summarization of Textbook Views

According to research by Willson (1982), from 1950 until the 1970's the research design statistics instruction for education graduate students stressed analysis of variance (ANOVA) techniques based on textbooks by Lindquist (1953), Winer (1962, 1971), Kirk (1969), Glass and Stanley (1970) and others. Not until the appearance of texts by Ward and Jennings (1973) and Kerlinger and Pedhazur (1973) could much emphasis on regression approaches be found. Willson (1982) attributes to these later texts the increased use of regression statistical approaches in OVA situations, as indicated by a review (Willson, 1980) of 10 years' research in the American Educational Research Journal. Only limited use of regression approaches was recorded from 1969 to 1978. In 1978-1979 regression methods were used extensively, according to Willson (1982).

An evaluation of five organizational behavior research journals from 1976 to 1985 was conducted by Gaither and Glorfeld (1985) with regression/correlation statistical analysis method being the most frequently employed technique. This shows the same trends in organizational behavior journals that is being documented in education journals, i.e., an increased use of regression methods.

A compendium of textbook views on planned versus post hoc tests was compiled from selected books to determine textbook views and recommendations regarding the use of planned versus post hoc multiple comparisons of means. Results of this compilation, along with selected quotations, are presented in Table 4.

INSERT TABLE 4 ABOUT HERE

Each book in Table 4 was randomly pulled and checked to see if planned or post hoc comparisons was indexed. Alternately, a reference for multiple comparisons of means was sought. From 1960-1967 no references were indexed for planned or post hoc comparisons (Armore, 1966; Bloomers & Lindquist, 1960; Bryant, 1960; Chase, 1967). Although Clark and Schkade (1969) do not cover planned versus post hoc comparisons, Kirk does in his 1969 book and explains that "techniques that have been developed for data snooping...are referred to as a posteriori or post-hoc tests" (p. 73). Glass and Stanley (1970) and Winer (1971), however, cover only post hoc comparisons in their texts. These findings corroborate Willson's (1980, 1982) view that statistics instruction did not historically offer the alternative planned comparison procedure.

Keppel (1973) took the stand that "post-hoc comparisons may be described as data 'sifting' (p. 93), intensive 'milking' (p. 133), and systematic 'fishing' (p. 136)." He stated that "it is clear, then, that tests of planned comparisons are a desirable alternative to the omnibus F test" (p. 93). This corresponds to Willson's findings (1982) that 1973 texts represented the beginning of emphasis on regression approaches (Kerlinger & Pedhazur, 1973; Ward & Jennings, 1973) and might further substantiate Willson's claim that these texts presenting planned comparisons helped contribute to the increased use of regression approaches in journal articles. Contrary to Willson's study, the present

research was not limited to textbooks used expressly in education graduate programs, but rather was a sampling across disciplines.

The use of planned versus post hoc comparisons continues to be a much debated issue with proponents of each view taking strongly felt and heatedly argued positions. But little or no emphasis was given this topic in many textbooks of the 1970's (Bhattacharyya & Johnson, 1977; Chase, 1976; Chiswick & Chiswick, 1975; Clark & Schkade, 1974; Glass & Stanley, 1970; Huck, Cormier, & Bounds, 1974; Jacobson, 1976; Mitroff & Kilmann, 1978; Nachmias & Nachmias, 1976; Winer, 1971). Even though Cohen and Cohen (1975) called "planned comparisons...the most elegant multiple comparison procedure [with] good power characteristics," they cautioned it could "only infrequently be employed in behavioral science investigations because the questions to be put to the data are simply not usually independent" (p. 158). Rosnow and Rosenthal (1989, p. 1282) credit current statistical textbooks with describing "the logic and the machinery of contrast analysis," but recognize that "one still sees contrasts employed all too rarely. That is a real pity given the precision of thought and theory they encourage and (especially relevant to these times of publication pressure) given the boost in power conferred with the resulting increase in .05 asterisks."

Statistical analysis is fast becoming a computerized automated process, allowing for faster, more accurate analyses of data. Statistical program documentation such as Norusis (1988a, 1988b), however, offers little conceptual framework for the student to become acquainted with

planned versus post hoc tests. This debate is one of substance and must not be overlooked in the classroom, or the researchers of tomorrow will lack important knowledge to design sound studies.

Recommendations

The present paper calls for increased emphasis on planned versus post hoc tests in the graduate student's statistics program. In addition, there is a need for further investigation of current statistics textbooks being used in the different schools of education, business, and the behavioral sciences to determine if the issue of planned versus post hoc tests is being adequately addressed in instructional materials.

A final note, echoing Rosnow and Rosenthal (1989, p. 1282), is that: "much of what we have said has been said before, but it is important that our graduate students hear it all again so that the next generation of psychological scientists is aware of the existence of these pitfalls and of the ways around them."

References

Abelson, R. P. (1962). Testing a priori hypotheses in the analysis of variance. Unpublished manuscript, Yale University, New Haven, CT. In R. L. Rosnow & R. Rosenthall, Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44(10), 1276-1284.

Armore, S. J. (1966). Introduction to statistical analysis and inference for psychology. New York, NY: John Wiley & Sons, Inc.

Bhattacharyya, G. K., & Johnson, R. A. (1977). Statistical concepts and methods. New York, NY: John Wiley & Sons.

Bloomers, P., & Lindquist, E. F. (1960). Elementary statistical methods in psychology and education. Boston, MA: The Riverside Press Cambridge.

Brite, R. L. (1983). Introduction to business statistics. Reading, MA: Addison-Wesley Publishing Company.

Bryant, E. C. (1960). Statistical analysis. New York, NY: McGraw-Hill Book Company, Inc.

Carver, R. P. (1978). The case against statistical significance testing. Harvard Educational Review, 48(3), 378-399.

Chase, C. I. (1967). Elementary statistical procedures. New York, NY: McGraw-Hill, Inc.

Chase, C. I. (1976). Elementary statistical procedures. (2nd ed.). New York, NY: McGraw-Hill, Inc.

Chiswick, B. R., & Chiswick, S. J. (1975). Statistics and econometrics: A problem-solving text. Baltimore, MD: University Park Press.

Clark, C. T., & Schkade, L. L. (1969). Statistical analysis for administrative decisions. Cincinnati, OH: South-Western Publishing Company.

Clark, C. T., & Lawrence, L. S. (1974). Statistical analysis for administrative decisions (2nd ed.). Cincinnati, OH: South-Western Publishing Company.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Daniel, L. G. (1989, January). Commonality analysis with various parametric methods. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX.

DuRapau, T. M. (1988, November). Benefits of using planned comparisons rather than post hoc tests: A brief review with examples. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (Eric Document Reproduction Service No. Ed 303 490)

Edwards, A. L. (1979). Multiple regression and the analysis of variance and covariance. San Francisco, CA: Freeman.

Elmore, P. B., & Woehlke, P. L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researchers, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Fish, L. J. (1988). Why multivariate methods are usually vital. Measurement and evaluation in Counseling and Development, 21, 130-137.

Fisher, R. A. (1925). Statistical methods for research workers. Edinburgh, England: Oliver and Boyd.

Fisher. R. A., & Yates, F. (1957). Statistical tables for biological, agricutural, and medical research (5th ed.). New York, NY: Hafner Publishing Co.

Gaither, N., & Glorfeld, L. (1985). An evaluation of the use of tests of significance in organization behavior research. Academy of Management Review, 10, 787-793.

Glasnapp, D. R., & Poggio, J. P. (1985). Essentials of statistical analysis for the behavioral sciences. Columbus, OH: Merrill.

Glass, G. V., & Stanley, J. C. (1970). Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice-Hall.

Goodwin, L. D., & Goodwin, W. L. (1985a). An analysis of statistical techniques used in the Journal of Educational Psychology, 1979-1983. Educational Psychologist, 20(1), 13-21.

Goodwin, L. D., & Goodwin, W. L. (1985b). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read educational research literature. Educational Researcher, 14(2), 3-11.

Gravetter, F. J., & Wallnau, L. B. (1985). Statistics for the behavioral sciences. St. Paul, MN: West.

Hays, W. L. (1981). Statistics (3rd ed.). New York, NY: Holt, Reinhart and Winston.

Hicks, C. R. (1973). Fundamental concepts in the design of experiments. New York, NY: Holt, Rinehard and Winston.

Huberty, C. J. & Morris, J. D. (1988). A single contrast test procedure. Educational and Psychological Measurement. 48, 567-578.

Huck, S. W., Cormier, W. H., & Bounds, Jr., W. G. (1974). Reading statistics and research. New York, NY: Harper and Row.

Jacobson, P. E., Jr. (1976). Introduction to statistical measure for the social and behavioral sciences. Hinsdale, IL: The Dryden Press.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Keppel, G. (1982). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

Kirk, R. E. (1969). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole.

Kirk, R. E. (1984). Elementary statistics. Monterey, CA: Brooks/Cole.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York, NY: Holt, Rinehart and Winston.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York, NY: Holt, Rinehart and Winston.

Lindquist, (1953). In V. L. Willson (1982, January). Misuses of regression approaches to ANOVA and ANCOVA. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX.

Minium, W. W., & Clarke, R. B. (1982). Elements of statistical reasoning. New York, NY: John Wiley and Sons.

Mitroff, I. I., & Kilmann, R. H. (1978). Methodoligical approaches to social science. San Francisco, CA: Jossey Bass, Inc., Publishers.

Moore, G. W. (1983). Developing and evaluating educational research. Boston, MA: Little, Brown and Company.

Nachmias, D., & Nachmias, C. (1976). Research methods in social sciences. New York, NY: St. Martin's Press.

Norusis, M. J. (1988a). The SPSS guide to data analysis for SPSS*. Chicago, IL: SPSS, Inc.

Norusis, M. J. (1988b). SPSS-X introductory statistics guide. Chicago, IL: SPSS, Inc.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York, NY: Holt, Rinehart and Winston.

Rosnow, R. L. & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44(10), 1276-1284.

Shavelson, R. J. (1988). Statistical reasoning for the behavioral sciences (2nd ed.). Boston, MA: Allyn & Bacon, Inc.

Snodgrass, J. G., Levy-Berger, G., & Haydon, M. (1985). Human experimental psychology. New York, NY: Oxford University Press.

Swaminathan, H. (1989). Interpreting the results of multivariate analysis of variance. In B. Thompson (Ed.), Advances in social science methodology (Vol. 1, pp. 204-232). Greenwich, CT: JAI Press, Inc.

Thompson, B. (1981, November). The problem of ovaism. Paper presented at the annual meeting of the Mid-South Educational Research Association, Lexington, KY.

Thompson, B. (1985). Alternate methods for analyzing data from education experiments. Journal of Experimental Education, 54, 50-55.

Thompson, B. (1986). ANOVA versus regression analysis of ATI designs: An empirical investigation. Educational and Psychological Measurement, 46, 917-928.

Thompson, B. (1987, November). The importance of a priori contrasts in analysis of variance research. Paper presented at the annual meeting of the Mid-South Educational Research Association, Mobile, AL. (Eric Document Reproduction Service No. ED 292 835)

Thompson, B. (1988a). A note about significant testing. Measurement and Evaluation in Counseling and Development, 20, 146-147.

Thompson, B. (1988b, November). Commom methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (Eric Document Reproduction Service No. ED 301 595)

Thompson, B. (1988c). The importance of planned or focused comparisons in OVA research. Measurement and Evaluation in Counseling and Development, 21, 99-101.

Thompson, B. (1989, January). Heuristics for understanding the concepts of interaction, polynomial trend, and the general linear model. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX. (Eric Document Reproduction Service No. Ed 304 453)

Ward, J. H., & Jennings, E. (1973). Introduction to linear models. Englewood, CA: Prentice-Hall, Inc.

Willson, V. L. (1982, January). Misuses of regression approaches to ANOVA and ANCOVA. Paper presented at the annual meeting of the Soutwest Educational Research Association, Dallas, TX.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.

Winer, B. J. (1962). Statistical principles in experimental design New York, NY: McGraw-Hill.

Winer, B. J. (1971). Statistical principles in experimental design. (2nd ed.). New York, NY: McGraw-Hill.

Table 1. Hypothetical Data for Attitudes Toward School Study (n=12)

ID	GROUP	LEVEL	DV	C1	C2	C3	C4	C5
1	Students	1	10	-1	-1	-1	-1	-1
2			20	-1	-1	-1	-1	-1
3	Teacher Aides	2	10	1	-1	-1	-1	-1
4			20	1	-1	-1	-1	-1
5	Teachers	3	10	0	2	-1	-1	-1
6			20	0	2	-1	-1	-1
7	Principals	4	10	0	0	3	-1	-1
8			20	0	0	3	-1	-1
9	Superintendents	5	10	0	0	0	4	-1
10			20	0	0	0	4	-1
11	Board Members	6	25	0	0	0	0	5
12			35	0	0	0	0	5

Table 2. One-Way ANOVA Results

SOURCE	SOS	df	MEAN SQUARE	F	p	ETA SQUARE
Between	375.000	5	75.000	1.500	.3155	.55556
Error	300.000	6	50.000			
Total	675.000	11				

NOTE: If, in violation of conventional practice, post hoc tests were conducted in the absence of a significant omnibus result, the tests would still not identify any two groups with means significantly different at alpha equal 0.05.

Table 3. Planned Comparison Results

CONTRAST SOURCE	SOS	df	MEAN SQUARE	F	p	ETA SQUARE
C1	.000	1	.000	0.000		.0000
C2	.000	1	.000	0.000		.0000
C3	.000	1	.000	0.000		.0000
C4	.000	1	.000	0.000		.0000
C5	375.000	1	375.000	12.500	.0054	.5556
Error	300.000	6	50.000			
Total	675.000	11				

SOURCE: Tables 1, 2, and 3 from "The Importance of a priori Contrasts in Analysis of Variance Research," Thompson (1987).

Table 4. Textbook Views on Planned Versus Post Hoc Tests

DATE	AUTHOR	CONTAINS PLANNED	CONTAINS POST HOC	QUOTATIONS
1960	BLOOMERS & LINDQUIST	NO	NO	
1960	BRYANT	NO	NO	
1966	ARMORE	NO	NO	
1967	CHASE	NO	NO	
1969	CLARK & SCHKADE	NO	NO	
1969	KIRK	YES	YES	Techniques that have been developed for data snooping...are referred to as a posteriori or post-hoc tests. (p. 73)
1970	GLASS & STANLEY	NO	YES	Most multiple comparison procedures are designed to be used after the null hypothesis of no treatment differences in an ANOVA has been rejected. (p. 383)
1971	WINER	NO	YES	Tukey and Scheffe developed methods for constructing simultaneous confidence intervals which avoid the pitfall of permitting the Type I error to become excessively large. (p. 175)
1973	KEPPEL	YES	YES	...a significant F allows, if not demands, a further analysis of the data. (p. 87) ...It is clear, then, that tests of planned comparisons are a desirable alternative to the omnibus F test. (p. 93) ...post-hoc comparisons may be described as "data sifting," where an experimenter is sorting through a large number of comparisons in the hope of finding something significant. (p. 93) ...post-hoc comparisons often take the form of an intensive "milking" of a set of results. (p. 133) ...This systematic "fishing" from the "pool" of

				all possible pairwise comparisons greatly increases the number of comparisons we are effectively considering and, therefore, the probability that we will make more Type I errors. (p. 136)
1974	CLARK & LAWRENCE	NO	NO	
1974	HUCK, CORMIER, & BOUNDS	NO	YES	...a researcher cannot stop his analysis after getting a significant F; he must locate the cause of the significant F. (p. 68)
1975	CHISWICK & CHISWICK	NO	NO	...regression analysis now tends to be preferred to ANOVA, except when tests for "any interaction" between two explanatory variables are of primary importance. (p. 211)
1975	COHEN & COHEN	YES	YES	Planned comparisons are generally considered the most elegant multiple comparison procedure and have good power characteristics, but...can only infrequently be employed in behavioral science investigations because the questions to be put to the data are simply not usually independent... (p. 158) ..."data snooping" is an important part of the research process, but unless Type I error is controlled...the experimentwise rate of spuriously significant t values on comparisons becomes unacceptably high. (p. 158)
1976	CHASE	YES	YES	Such "data snooping" is known as post hoc comparisons...(p. 188)
1976	JACOBSON	NO	NO	
1976	NACHMIAS & NACHMIAS	NO	NO	
1977	BHATTACHARYYA & JOHNSON	NO	NO	
1978	MITROFF & KILMANN	NO	NO	

1979	EDWARDS	YES	NO	...there are many other ways in which mutually orthogonal X vectors can be constructed...some sets may be more meaningful than others in a given experiment. (p. 85)
1983	BRITE	NO	NO	
1985	SNODGRASS, LEVY-BURGER, & HAYDON	YES	YES	...it is possible that a particular planned comparison will be significant in the absence of an overall significant F from the ANOVA. (p. 386)
1988a	NORUSIS	NO	NO	In most situations, you want to pinpoint exactly where the differences are. To do this, you must use multiple comparison procedures...(pp.262-263)
1988b	NORUSIS	NO	NO	A variety of special techniques, termed multiple comparison procedures, are available for determining which population means are different from each other...(p. 119)
1988	SHAVELSON	YES	YES	...planned comparisons are more powerful than the data snooping permitted by post hoc comparisons. (p. 364)

APPENDIX A
Selected SPSS-X Control Cards

```
TITLE '*****OMNIBUS no   POSTHOC  no   A PRIORI  yes'
FILE HANDLE BT/NAME='APRIORI.DTA'
DATA LIST FILE=BT/LEV 1 DV 2-4
COMPUTE C1=0
COMPUTE C2=0
COMPUTE C3=0
COMPUTE C4=0
COMPUTE C5=0
IF (LEV EQ 1)C1=-1
IF (LEV EQ 2)C1=1
IF (LEV EQ 3)C2=2
IF (LEV EQ 1 OR LEV EQ 2)C2=-1
IF (LEV EQ 4)C3=3
IF (LEV LT 4)C3=-1
IF (LEV EQ 5)C4=4
IF (LEV LE 4)C4=-1
IF (LEV EQ 6)C5=5
IF (C5 EQ 0)C5=-1
REGRESSION VARIABLES=DV C1 TO C5/DESCRIPTIVES=ALL/
  CRITERIA=PIN(.95) POUT(.999) TOLERANCE(.00001)/DEPENDENT=DV/
  ENTER C5/ENTER C4/ENTER C3/ENTER C2/ENTER C1/
```

Note: Adapted from "The importance of a priori contrasts in analysis of variance research: (Thompson, 1987).